DOCUMENT RESUME

ED 235 228 TM 830 650

AUTHOR	Rogers, Bruce G.
TITLE	A Time Series Approach to the Longitudinal Study of Undergraduate Grades.
PUB DATE	13 Apr 83
NOTE	19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, April 12-14, 1983).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Grade Inflation; Grade Point Average; *Higher Education; *Longitudinal Studies; *Mathematical Models; Research Methodology; Undergraduate Students
IDENTIFIERS	*Time Series Analysis; Time Series Design

ABSTRACT

During the past 15 years, considerable attention has been given to a conspicuous longitudinal change in grading patterns in higher education. Commonly referred to as "grade inflation," the phenomenon has been perceived by some as seriously weakening the meaning of grades but by others as reflecting a positive tendency for students to select those courses which reflect their own ability. Researchers have continued to study the trends, but no models have yet been presented which explain the variations, across an extended time period, in terms of measurable independent variables. It was the purpose of this study to search for an explanatory model of longitudinal collegiate grade point average (GPA) data, using a time series methodological approach. Fifty-three years of GPA data from a midwestern university were analyzed with exogenous variables (changes in student ability, enrollment level, sex distribution, and the national economic status) and endogenous lagged variables. Only the enrollment level was found capable of explaining any significant part of grade variation. It was concluded that short term ex post facto studies may have limited value unless independent variables can be shown to have longitudinal explanatory power. (Author/PN)

A TIME SERIES APPROACH TO THE LONGITUDINAL STUDY OF UNDERGRADUATE GRADES

Bruce G. Rogers

Department of Educational Psychology
University of Northern Iowa
Cedar Falls, Iowa

Paper presented at the annual meeting of the National Council of Measurement in Education, April 13, 1983, Montreal, Quebec, Canada.

Abstract

A Time Series Approach to the Longitudinal Study of Undergraduate Grades

Bruce G. Rogers, University of Northern Iowa

It was the purpose of this study to search for an explanatory model of longitudinal collegiate grade point average data, using a time series methodological approach. Fifty three years of GPA data from a midwestern university were analyzed with exogenous variables (changes in student ability, enrollment level, sex distribution, and the national economic status) and endogenous lagged variables. Only the enrollment level was found capable of explaining any significant part of grade variation. It was concluded that short term ex-post facto studies may have limited value unless independent variables can be shown to have longitudinal explanatory power.

A TIME SERIES APPROACH TO THE LONGITUDINAL STUDY OF UNDERGRADUATE GRADES

During the past fifteen years, considerable attention has been given to a conspicuous longitudinal change in grading patterns in higher education. Commonly referred to as "grade inflation," the phenomenon has been perceived by some as seriously weakening the meaning of grades (Juola, 1979) but by others as reflecting a positive tendency for students to select those courses which reflect their own ability (Prather, et. al., 1979). Although the upward movement has subsequently abated (Suslow, 1977), grades remain at a considerably higher level than that prior to the onset of "grade inflation." Researchers have continued to study the trends, but no models have yet been presented which explain the variations, across an extended time period, in terms of measurable independent variables.

In the statistical literature, the investigation of temporal sequences is referred to as a time-series analysis; it is commonly used by econometricians to explore fluctuations in economic data. Applications in education have most frequently used the interrupted time series design (Cook and Campbell, 1979), but in the analysis of grade inflation, no identifiable interruption point exists, so the appropriate model appears to be a noninterrupted design.

Factors Affecting Grades

A review of the extant literature indicates four general categories of variables which have been suggested as being causally related to grade inflation.

1. Student ability: It would seem reasonable to begin with the assumption that a grade should reflect the competence of the pupil, and thus observed variations in grades should reflect observed variations in competence. Aiken (1963) and Wilson (1970) investigated this hypothesis using data from the early and middle 1960's. They each found that measures of ability of successive freshmen classes tended to increase during this period, but without a concomitant rise in grade point average. Birnbaum (1977) found that in the late 1960's and early 1970's, a different, but still discomfirming trend occured when grade point averages showed a marked increase, but without a corresponding increase in student ability. Birnbaum also found that there was no corresponding increase in student achievement as measured by the Graduate Record Examination. These results, however, need to be replicated over longer time spans.

2. Sex ratio: Since females tend to obtain higher grades than males (Schroeder and Sledge, 1966), changes in the ratio of females to males may be thought to explain grade variations. Although Birnbaum (1977) failed to find support for this hypothesis, the results need to be replicated in other settings before the hypothesis can be discarded.

3. Internal institutional changes: A number of institutional changes have been hypothesized as causes of grade variation, including grading practices (e.g., Pass/Fail grades), student evaluations of faculty, changes in the types of courses selected by students, and fluctuations in enrollment. Some empirical studies have been conducted, but the interpretation of such data analyses is problematical, since the results are not consistent across studies. Indeed, each of these changes, and others, occurred to some extent during the years usually associated with grade inflation, but the data do not permit conclusions upon which researchers have concurred (e.g., Prather, et. al., 1979; Carney, et. al., 1978; Birnbaum, 1977).

4. Societal factors: The process of grading takes place in environments that are effected by social forces, such as economic changes, wars, and civil rights movements, all of which have been suggested as affecting the way grades are assigned. Although there appears to be little or no published empirical data on these relationships, the importance that is put upon them (e.g., Birnbaum 1977) would seem to warrant further investigation.

It was the purpose of this research to investigate the extent to which grade variations over time could be predicted by measurable indices from each of the above categories. The statistical techniques of time series analysis have heretofore been unreported in the literature pertaining to grade inflation. This paper reports an initial exploratory data analysis using the most basic concepts of time series analysis.

In its present state of the art, time series analysis contains a wide variety of techniques, ranging from rather simple regression analyses to sophisticated techniques for extensive data modeling. For this study, it was decided to use a relatively straightforward design, primarily for two reasons. First, time series analyses require relatively long series; at least 50 observations, according to one text on the subject (McCleary and Hay, 1980, p. 20). Since our data base only barely met that minimum requirement, it seemed prudent not to begin with an analysis whose demands exceeded what the data would support. Second, it seems a safe assumption that all of the members of this audience are already reasonably familiar with simple regression, and could apply these techniques to data in their own institutions. If that type of analysis shows promise, then there is justification for seeking more sophisticated analyses.

The perceptive reader will, no doubt, find many aspects of this study that could be improved; if so, a major purpose of the paper will have been accomplished by motivating others to seek even better models for explaining this phenomenon.

Method

Sample

The University of Northern Iowa has evolved, over the past century, from a State Normal School to a medium scope university of approximately 11,000 students. In 1929, the school adopted the now almost ubiquitous 5-point grading system. Figure 1 shows the grade point average (GPA) data of this institution for each year from 1929 through 1981. Thus, a total of 53 data points are available for analysis.

To determine if the grading trends in the data were comparable to those in other institutions, grade data was collected from two other universities; from University A since 1947 and from University B since 1958. Several surveys by Juola (1979) of institutions of higher education provided "national" data for selected years between 1960 and 1978. Figures 2, 3, and 4 show the comparison of these data with the UNI data. It appears that, at least for the past two decades, trends in the UNI data are reasonably comparable with other institutions of higher education.

Measures of the Variables

From each of the four classes of variables, it was desired to select a measure that would have data available for each of the past twenty years or more.

1. Student Ability. Beginning in 1959, all students applying to attend the school took the ACT tests; average composite scores were available for each year since then. This measure is generally regarded as one of the best indicators of student ability. Unfortunately, no suitable data on student ability were available for the years prior to 1959.

2. Sex Ratio. Values of the ratio of number of women to men were available for each of the 53 years covered by the study. This ratio, rather than its inverse, was chosen because for every year except one, the women outnumbered the men. Thus, in general, the index can be readily interpreted as reflecting the preponderance of women over men.

3. Internal Institutional Changes. In searching for a variable in this category which would yield data for each year of the study, it became apparent that the annual enrollment data were probably the most useful figures at our disposal for reflecting the evolution of the institution.

It should, perhaps, be noted that factors such as pass-fail grading, contract grading, student evaluations, etc., are often postulated in the literature as causative factors. However, since no time-series data were available for any of these measures, they were not included in this analysis. Further comments will be made on this topic in the Discussion Section.

4. Societal Factors. Of the many social forces which could potentially affect the University on an annual basis, the state of the economy is one of constant concern in terms of its effects on the composition of the student body. The widely used Consumer Price Index (CPI) was selected as an appropriate measure to represent this category (Bureau of the Census, 1975, 1981).

While some writers have suggested the Vietnam War draft, the affirmative action movement, etc. as causative factors, again the lack of annual time series data on these variables precluded their use in this analysis.

Procedures and Results

Because of the exploratory nature of the analysis, it will be more efficient to depart slightly from standard practice and present the procedure and results together.

The analysis began with an examination of student ability, as reflected by the average ACT scores. As shown in Figure 5, these scores tended to rise during the decade of the 1960's, but subsequently exhibited a marked decline. When the 23 ACT score points were plotted on a scattergram to predict Freshman GPA values, the resulting pattern actually resembled the letter C written backwards. If data from each year are analyzed separately, there is a moderately strong relationship between ACT and GPA within years, but the data across years failed to reveal any resemblance of a comparable association. It was, therefore, concluded that further time series analysis with this variable would not be fruitful.

The analysis continued with an examination of the economic, enrollment, and sex ratio variables as potential components of a predictive model for GPA. Each variable was a sequence of 53 numbers.

When examining any sequence, the statistician is likely to first ask if it reflects a stationary process, both in level and variance. A priori, it seems reasonable to assume that GPA should be stationary when considered over a very long period of time, since it is difficult to postulate any underlying force which would require it to increase, decrease, or fluctuate in any predictable manner. However, as is obvious by inspection from Figure 1, the sequence plot for GPA is not stationary, but shows both a general increase over time and strong sequencing effects. The other three variables likewise showed non-stationary effects, as indicated by the large autocorrelations shown in Table 1 for the data before transformation. Accordingly, the variables were subjected first to a logarithmic transformation and then to first differences, wherein each value was replaced by the difference between it and the preceding value (McCleary and Hay, 1980, pp. 40, 51). As also shown in Table 1, the transformations did not completely stabilize the behavior of these variables, but they did improve it considerably, as shown by the reduction in autocorrelations and the changes in the Box-Pierce statistics.

The number of autocorrelations computed for investigating the behavior of data is often determined by an expected seasonality. Thus, if monthly data is expected to be influenced by annual seasonality, then at least twelve lags will be required to detect it. However, since the GPA data exhibited no evidence of anything like a seasonal pattern, that consideration was deemed irrelevant. Another consideration is that of sampling variability; a conservative recommendation is to limit the number of lags to about one-fifth of the sample size (Roberts, 1974, p. 164). It therefore appeared that the first ten autocorrelations would probably be most appropriate for this data.

To test the hypothesis that the economic variable (CPI), the enrollment, and the sex ratio would be useful in modeling GPA, a multiple regression analysis was conducted. Only the variable of enrollment showed statistical significance; the correlation coefficient was .36, with p = .008. The residuals of this model showed reasonably random behavior, as indicated by their small autocorrelations (see Table 1).

To investigate the possibility that curvilinear functions and interactions would improve the fit of the model, the squared terms and crossproducts of the variables were also tested. However, none produced increases in the multiple R that could be considered to be statistically significant, so it was concluded that they did not show evidence to warrant continued investigation on this data.

Lagged Variables

Time series data lend themselves to the analysis of lagged variables (Ostrom, 1978, p. 44 ff). Two sets of exogenous lagged variables were considered; the first four lags on enrollment and the first four lags on CPI. In both cases the question addressed was whether the lagged variables would add to the prediction given by the unlagged variable.

In the case of CPI, no combination of the lags produced any significant results. In the case of enrollment, the multiple correlation showed an increase from .36 to .43 when the third lag was entered. Such a small increase, however, might be just the result of a statistical aberation in the data, particularly when it is considered that the resulting p values increased from .008 to .03.

Perhaps there is what might be called a "grading momentum" wherein the change in the level of grades assigned in any given year is expected to be a function of the changes that occurred in previous years. Such predictors are termed lagged endogenous variables. The first four lags were used to test the hypothesis for this data, but none of the multiple correlation coefficients even approached statistical significance.

Discussion

This data anlysis began by comparing the sequence of yearly average ACT scores, with the sequence of GPA values. The comparison clearly revealed results contrary to the positive linear relationship that would be expected if changes in the average GPA were a result of changes in the average student ability. These results are thus consistent with those of previous investigators and suggest that hypotheses invoking changes in student ability appear not to be defensible explanations for the observed longitudinal changes in grading patterns (Baird and Flister, 1972).

The results also suggest that changes in grading patterns, in at least one university, seem to have been occurring concomitantly with changes in enrollment figures. However, this relationship was rather weak, and the resulting model certainly did not provide a strong fit to the data. The use of the grade variations of previous years also failed to produce an acceptable model.

What do we do when we fail to find a good model for data? If we assume that we have exhausted the plausible procedures for model building, then we perhaps can learn what we do not know about the problem at hand. It may make us a little cautious about those explanations that have been generated on data sets of short duration; for while they can explain the data in a post hoc analysis, they may fail to make accurate predictions of future outcomes. Such explanations have questionable value in any scientific structure.

It should be emphasized that this particular study was exploratory in its nature, and at least three of its limitations should be clearly kept in mind. First, it was done at only one university. While the grade inflation trends appeared similar to those at other universities, the enrollment patterns were not, and the observed relationship of the enrollment variable to grade changes may fail to be generalizable to other settings.

A second limitation of this study is that a very simple statistical analysis was used. In the class of ARIMA multivariate time series, there are much more sophisticated procedures than were used in this study, but this simple procedure was selected as most appropriate for a beginning exploratory study. This author intends to extend the analyses of these types of data with more elaborate models and hopes that others will apply them also.

A third limitation concerns the choice of measures for each category of variables. In particular, the CPI and Enrollment measures may not adequately reflect the effect of the economy and institutional changes upon the way grades are assigned. While these were the only measures of these variables that were examined, perhaps others would show more promise.

Finally, the study is limited in that there was no consideration of the impact of a discrete event, such as a particular war, a new policy, etc. While the interpretation of the results of such interrupted designs may be susceptible to the internal validity threat of history, since several other events may have occurred near the same time, it would still be of value to ascertain if the results of a time series analysis would lend support to the effect of such impacts.

In conclusion, it can be said that while it is easy to speculate on what factors have caused the grade inflation, it is more difficult to find factors which can be depended upon to explain grade fluctuations over the extended time period of the last half century. Using short term studies, it is always possible to find measures or events to explain any given grade fluctuation, but unless long-term predictive power is demonstrated, we probably should be a little suspect of their explanatory value. Clearly, the area warrants further study with a variety of data analysis techniques.

References

Aiken, L.R. The grading behavior of a college faculty. Educational and Psychological Measurement, 1963, 23, 319-22.

Baird, L. and Flister, A. Grading students: The relation of changes in average student ability to the average grades awarded. American Educational Research Journal, 1972, 9, 431-42.

Birnbaum, R. Factors relating to university grade inflation. Journal of Higher Education, 1977, 48, 519-539.

Bureau of the Census. Historical statistics of the U.S., Washington, D.C.: The Bureau. 1975.

Bureau of the Census. Statistical abstracts of the U.S., Washington, D.C.: The Bureau. 1981.

Carney, P., Isakson, R. L., and Ellsworth, R. An exploration of grade inflation and some related factors in higher education. College and University, 1978, 53, 217-230.

Juola, A. E. Grade inflation in higher education - 1979, is it over? East Lansing: Learning and Evaluation Service, Michigan State University, 1979.

McCleary, R. and Hay, R. A. Applied time series analysis for the social sciences. Beverly Hills: Sage Publications, 1980.

Ostrom, C. W. Time series analysis: Regression techniques. Sage Series No. 07-009. Beverly Hills: Sage Publications, 1978.

Prather, J. E.; Smith, G. , and Kodras, J. E. A longitudinal study of grades in 144 undergraduate courses. Research in Higher Education, 1979, 10, 11-24.

Roberts, H. V. Conversational Statistics. New York: McGraw Hill, 1974.

Schroeder, W. L. and Sledge, G. W. Factors related to collegiate academic success. The Journal of College Student Personnel, 1966, 7, 97-104.

Suslow, S. Grade inflation: End of a trend? Change, March 1977, 44-45.

Wilson, K. J. Increased selectivity and institutional grading standards. College and University, 1970, 46, 46-53.

Footnote

Appreciation is hereby expressed to Dr. Kenneth J. DeNault, Department of Geology, University of Northern Iowa, for the preparation of Figures 2, 4, and 5.

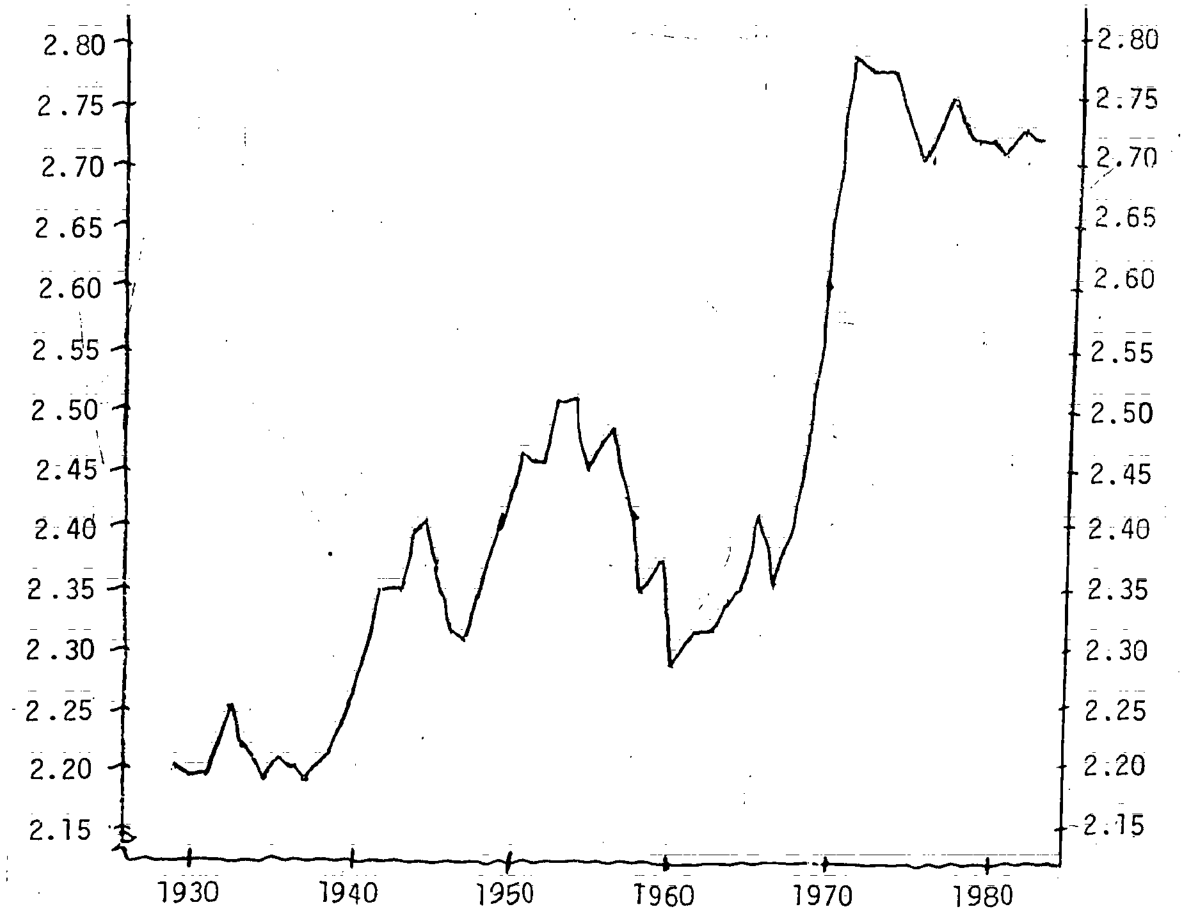

Figure 1. Grade point averages at UNI, by year, from 1929 to 1981. (Before 1944 is whole year data, afterward is fall term.)

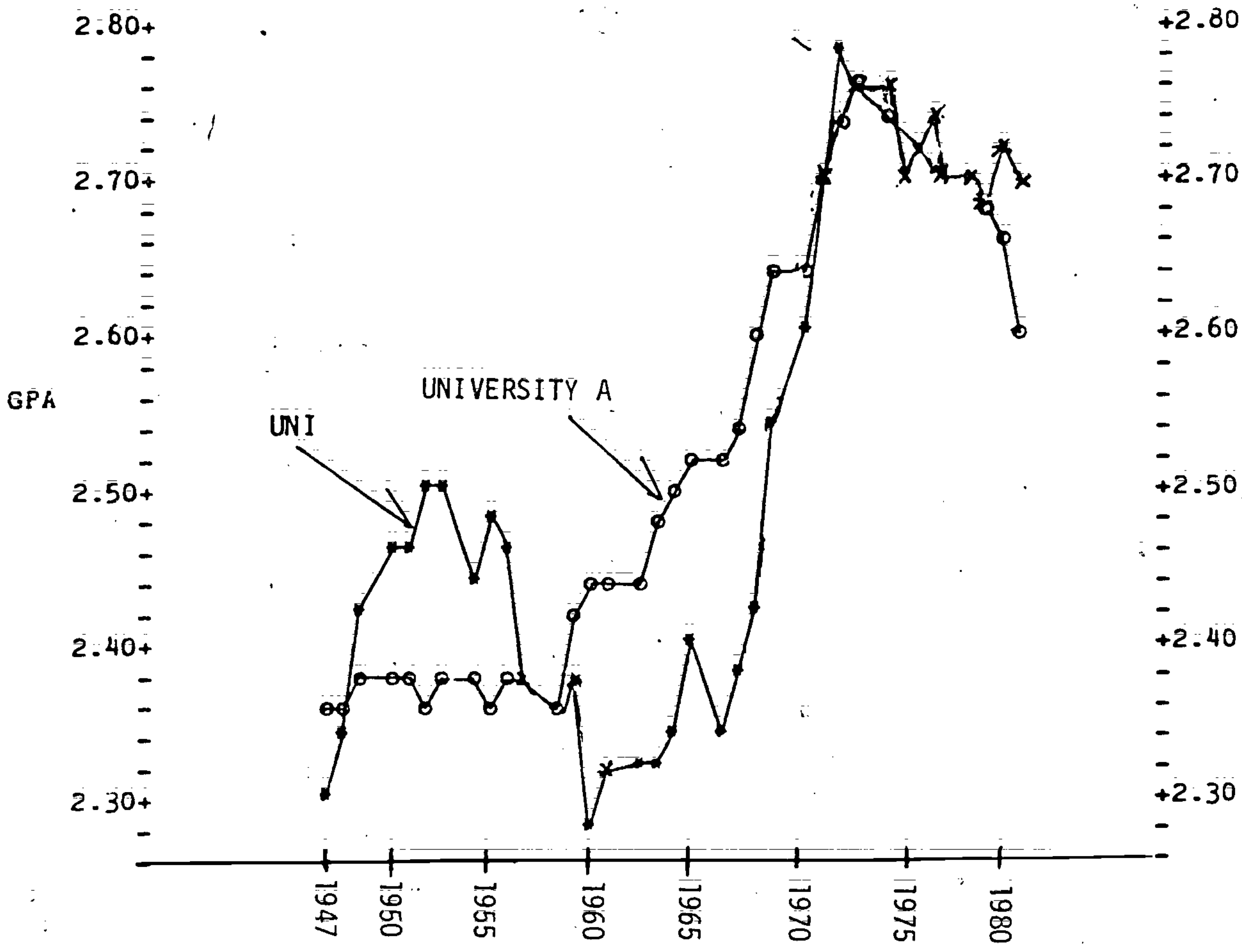

Figure 2. Average fall term undergraduate GPA at the University of Northern Iowa and University A (1947-1981).

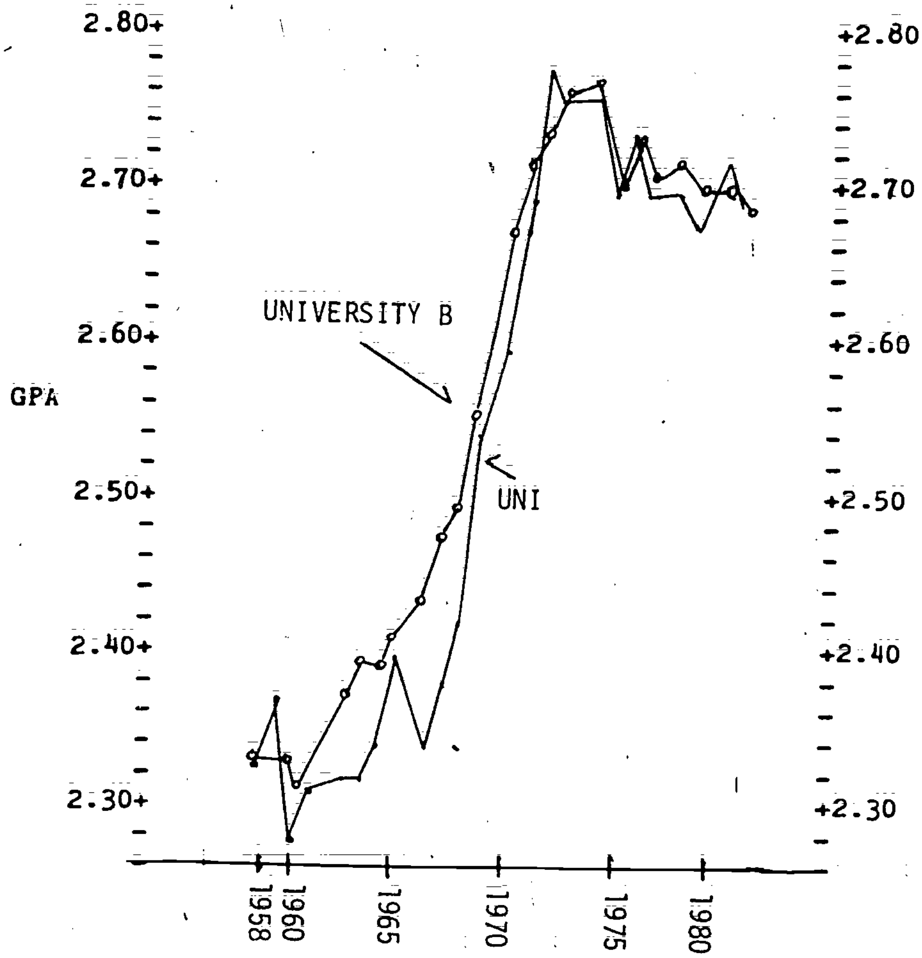

Figure 3. Average fall term undergraduate GPA at the University of Northern Iowa (UNI) and University B for 1958-1981.

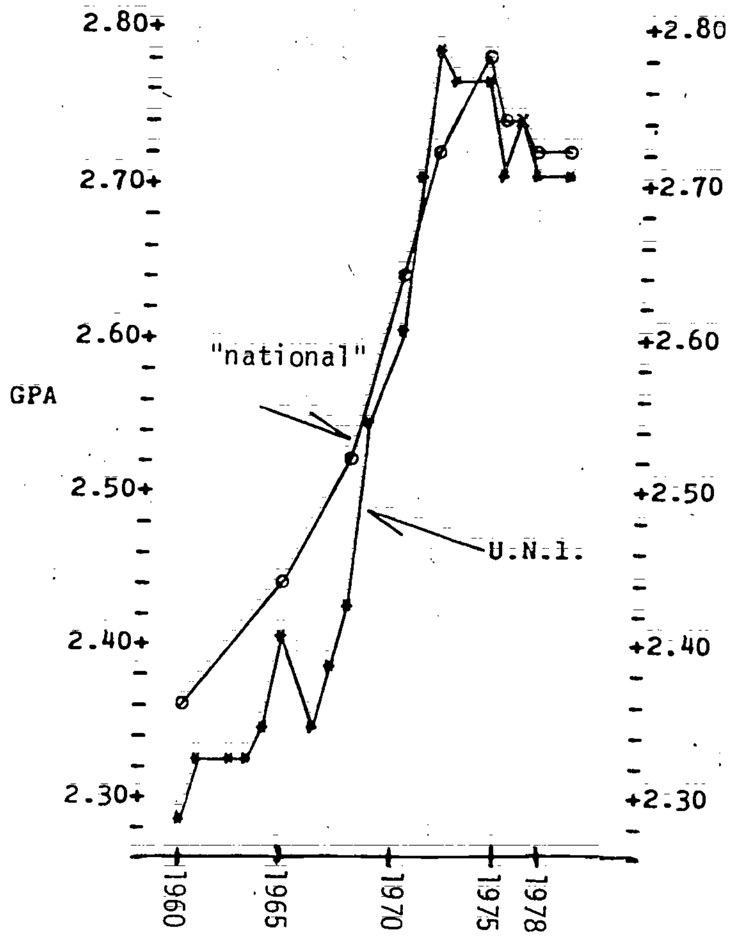

Figure 4. Average fall semester undergraduate GPA at the University of Northern Iowa (UNI) and "national" GPA (1960-1978).

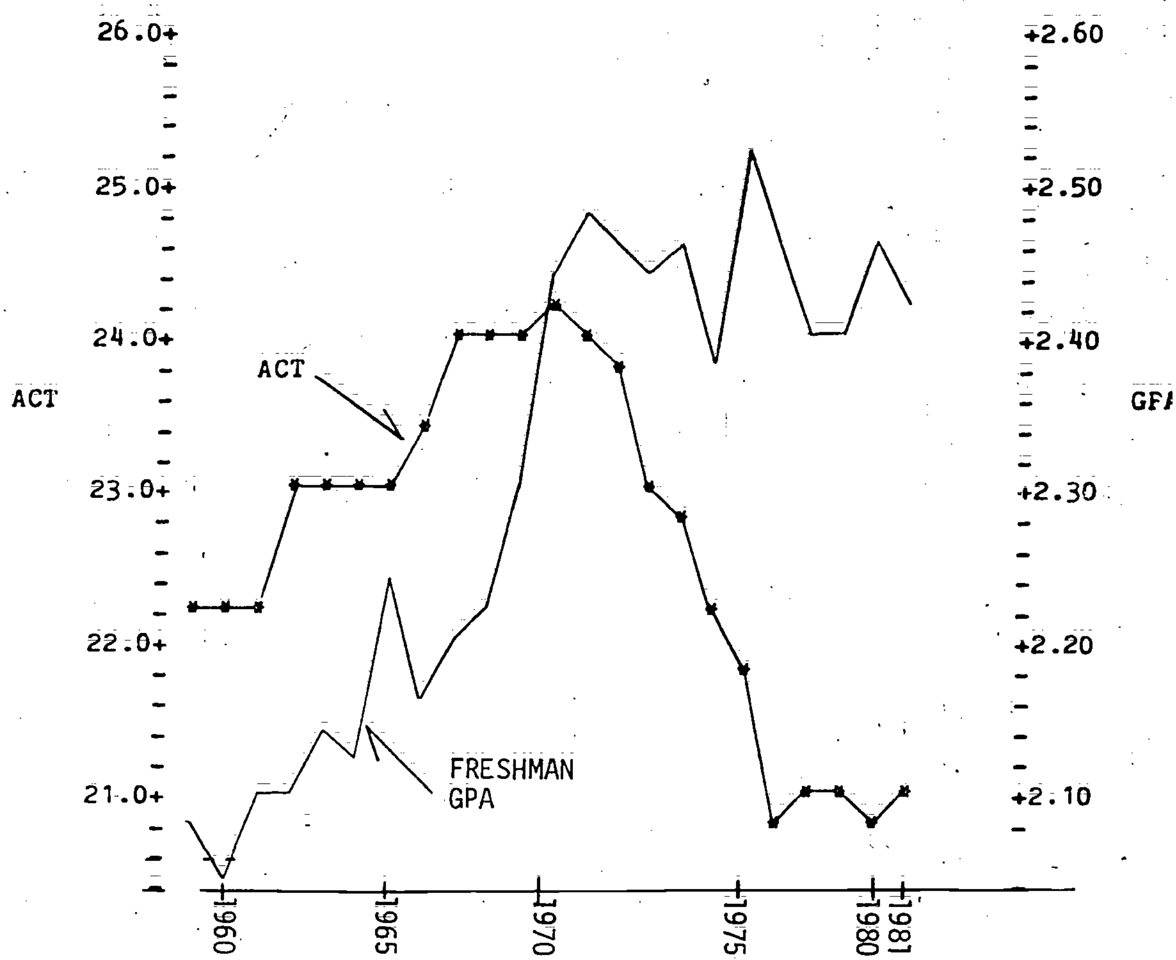

Figure 5. Freshmen at UNI: Fall semester mean GPA and mean ACT composite scores (1959-1981).

TABLE 1

Autocorrelations of variables before and after the log difference transformations.

Before Transformation		After Transformation	
GPA		GPA	
ORDER	AUTOCORR.	ORDER	AUTOCORR.
1	.93	1	.12
2	.84	2	.14
3	.75	3	.04
4	.65	4	.03
5	.55	5	-.33
6	.46	6	.00
7	.36	7	-.12
8	.27	8	-.06
9	.17	9	.08
10	.06	10	-.17
SE	.137	SE	.139
BP	174.621	BP	10.290
EXP	10.0	EXP	10.0
SNA	10.844	SNA	.213
CPI		CPI	
ORDER	AUTOCORR.	ORDER	AUTOCORR
1	.88	1	.70
2	.77	2	.28
3	.68	3	.15
4	.60	4	.23
5	.53	5	.32
6	.46	6	.24
7	.39	7	.09
8	.34	8	-.01
9	.29	9	-.05
10	.25	10	-.10
SE	.137	SE	.139
BP	163.501	BP	42.655
EXP	10.0	EXP	10.0
SNA	10.466	SNA	4.320

Note:
SE = Standard error of each coefficient, given the random model.
BP = Box-Pierce Statistic
EXP = Expected value, given a random model.
SNA = Standard Normal Approximation

TABLE 1 (Continued)

Before Transformation		After Transformation	
Enrollment		Enrollment	
ORDER	AUTOCORR.	ORDER	AUTOCORR.
1	.95	1	.37
2	.90	2	.11
3	.84	3	.35
4	.78	4	.23
5	.71	5	.29
6	.66	6	.15
7	.61	7	.01
8	.56	8	.13
9	.51	9	.22
10	.46	10	.21
SE	.137	SE	.139
BP	271.438	BP	28.186
EXP	10.0	EXP	10.0
SNA	13.601	SNA	2.916

Sex Ratio		Sex Ratio	
ORDER	AUTOCORR.	ORDER	AUTOCORR.
1	.68	1	.25
2	.23	2	-.15
3	-.01	3	-.50
4	-.02	4	-.14
5	-.01	5	-.12
6	.03	6	-.00
7	.06	7	.03
8	.08	8	.14
9	.08	9	.03
10	.06	10	.03
SE	.137	SE	.139
BP	28.244	BP	20.464
EXP	10.0	EXP	10.0
SNA	2.923	SNA	1.957

TABLE 2

Autocorrelations on the residuals for the regression of Enrollment on GPA.

ORDER	AUTOCORR.
1	.05
2	.18
3	.12
4	.14
5	-.25
6	.10
7	-.12
8	-.15
9	.02
10	-.23
SE	.139
BP	11.826
EXP	10.0
SNA	.534

Note: See Note on Table 1.